A computational algebraic geometry package for graphical models

Alexander Diaz, Luis David Garcia-Puente, Shaowei Lin,
Sonja Petrović, Michael Stillman

Abstract

We introduce a `Macaulay2` package `GraphicalModels.m2`, which constructs ideals and polynomial rings corresponding to discrete and Gaussian graphical models from the point of view of algebraic statistics. The main functions of the package compute conditional independence statements implied by the graph. One of the applications of the package is in solving the parameter identifiability problem for Gaussian graphical models.

1 Introduction

Algebraic statistics uses algebraic geometry and related fields for the study of statistical models. The computational nature of many fundamental questions in the field shows the need for computational algebraic methods tailored specifically to algebraic statistics. The study of graphical models plays an important role in the field ([7], [2], **some more citations here**), and with its rich (and recent) history is in itself deserving of a package.

The core of the package was developed by the second and fifth authors and is based on the constructions described in [4]. These methods are used to construct ideals corresponding to discrete graphical models. Following [9] and [10], we also implement these constructions for Gaussian Bayesian networks and for graphs containing both directed and bidirected edges.

Further, the package contains methods that can be used to solve the identifiability problem for Gaussian graphical models, as described in [11]. As the research on algebraic statistics for graphical models continue, we envision the package growing to include more methods and interface with other packages, such as `NumericalAlgebraicGeometry`. **DO WE want to say this? Anton's package paper isn't published yet..**

We use the `Graphs` package to create graphs for the models.

2 Mathematical Background

For the purpose of this note and the package, an algebraic parametric statistical **model** is a set of probability distributions parametrized by a given polynomial map. Since probabilities are nonnegative and sum to 1, each model is obtained by intersecting the corresponding variety (that is, the Zariski closure of this map considered over the complex numbers) with the probability simplex. For graphical models, the defining ideals of these varieties can be obtained from the graph. In what follows, we explain how this is done.

2.1 Discrete Graphical Models

Let $X = \{X_1, \ldots, X_n\}$ be a collection of discrete random variables such that X_i takes values in the finite set $[d_i] = \{1, 2, \ldots, d_i\}$. By $\mathrm{Prob}(X = x)$ we will denote the joint probability $\mathrm{Prob}(X_1 = x_1, \ldots, X_n = x_n)$ of observing the variable X_i in state $x_i \in [d_i]$. Let $p_{x_1 \cdots x_n}$ be an indeterminate representing $\mathrm{Prob}(X = x)$, and consider the polynomial ring $\mathbb{R}[p]$ in $d_1 \cdots d_n$ indeterminates.

For pairwise disjoint subsets $A, B \neq \emptyset$, and C of $\{X_1, \ldots, X_n\}$, we say that A is independent of B given C, written $A \perp\!\!\!\perp B \mid C$, if

$$\begin{aligned} &\mathrm{Prob}(A = a, B = b, C = c) \cdot \mathrm{Prob}(A = a', B = b', C = c) \\ = &\mathrm{Prob}(A = a, B = b', C = c) \cdot \mathrm{Prob}(A = a', B = b, C = c) \end{aligned} \tag{1}$$

for all values a, a', b, b', c which the random variables in A, B, C can take. By marginalizing (???) over all values of random variables not in $A \cup B \cup C$, each of the joint probabilities above can be expanded as a sum of indeterminates $p_{i_1 i_2 \cdots i_n}$. Thus, conditional independence statements translate into homogeneous quadratic polynomials in $\mathbb{R}[p]$.

Given a directed acyclic graph G on vertices $\{1, \ldots, n\}$, let $\mathrm{pa}(i)$ denote the set of parents of a vertex i, and $\mathrm{nd}(i)$ the set of its nondescendents. We say that X factors according to G if

$$\mathrm{Prob}(X = x) = \prod_{i=1}^{n} \mathrm{Prob}(X_i = x_i \mid X_{\mathrm{pa}(i)} = x_{\mathrm{pa}(i)}).$$

for all values $x \in D$. It can be shown that if X factors according to G, then the following conditional independence statements are satisfied:

1. Pairwise Markov. If $i \to j$ is not an edge of G, then $X_i \perp\!\!\!\perp X_j \mid X_{\mathrm{nd}(i) \setminus j}$.

2. Local Markov. For all vertices i of G, then $X_i \perp\!\!\!\perp X_{\mathrm{nd}(i)} \mid X_{\mathrm{pa}(i)}$.

3. Global Markov. If C d-separates A and B, then $A \perp\!\!\!\perp B \mid C$.

Here, we say that C d-separates (direct-separates) A and B in G if every path (not necessarily directed) from A to B contains either a noncollider $i \in C$ or a collider i with $C \subseteq \mathrm{nd}(i)$, where a vertex i is a collider if $a \to i \leftarrow b$ on the path.

Compute conditional independence statements using the Bayes ball algorithm. **do we simply state this here, or do we not state it at all??**

2.2 Gaussian Graphical Models

Consider a **mixed graph** G with vertices $V = \{1, \ldots, n\}$ and edges E which could be directed $i \to j$, undirected $i - j$ or bidirected $i \leftrightarrow j$. We assume that there is a partition $U \cup W = \{1, \ldots, k\} \cup \{k+1, \ldots, n\}$ of V such that all undirected edges have their vertices in U, all bidirected edges have their vertices in W, and there are no directed edges $w \to u$ with $w \in W$ and $u \in U$. **assume? is this a strong assumption? deserves a word...**

Let Λ be an $n \times n$ matrix with $\Lambda_{ij} = 0$ if $i \to j \notin E$ and $\Lambda_{ii} = 0$ for all i. Let K and Φ be symmetric positive definite matrices, with rows and columns indexed by U and by W respectively, such that $K_{ij} = 0$ if $i - j \notin E$, $\Phi_{ij} = 0$ if $i \leftrightarrow j \notin E$ and $K_{ii}, \Phi_{ii} > 0$ for all i. Now, consider normal random variables $X_1, \ldots, X_n$ with zero means and covariance matrix

$$(I - \Lambda)^{-T} \begin{pmatrix} K^{-1} & 0 \\ 0 & \Phi \end{pmatrix} (I - \Lambda)^{-1}.$$

Hence, the joint distribution of these random variables is parametrized by the $\Lambda_{ij}, K_{ij}, \Phi_{ij}$ corresponding to directed, undirected and bidirected edges of G, respectively, as well as the K_{ii}, Φ_{ii} corresponding to vertices in U and W.

TO DO : conditional independence coming from trek separation, they translate to rank conditions on Σ

regarding the next statement: I think we don't have to do this here; we can just show it on an example in the next section and put a reference to the theorem? Algorithm for computing using Theorem so and so in Seth's paper. Creating an auxiliary DAG.

2.3 Identifiability of Gaussian Graphical Models

A word on the identifiability problem. A fundamental question regarding any statistical model is to determine if the parameters that gave rise to the given data point can be recovered uniquely. **blabla** algebraically this translates to checking if the parametrization map is generically one-to-one.

3 Examples

Consider the "diamond" graph G: $d \to c$, $b \to a$:

```
i1 : G = digraph  {{a,{}},{b,{a}},{c,{a}},{d,{b,c}}}
o1 = Digraph{a => set {}    }
             b => set {a}
             c => set {a}
             d => set {b, c}
o1 : Digraph
```

The following code creates the ideal in 16 variables whose zero set represents the probability distributions on four binary random variables which satisfy the conditional independence statements coming from G. Each statement $A \perp\!\!\!\perp B \mid C$ is represented by a list $\{A, B, C\}$.

```
i2 : R = markovRing (2,2,2,2)
o2 = R
o2 : PolynomialRing
i3 : S = globalMarkovStmts G  -----RENAME THIS FUNCTION ALREADY?! STMTS!!
o3 = {{{a}, {d}, {b, c}}, {{b}, {c}, {d}}}
o3 : List
i4 : I = markovIdeal(R,G,S) -- NO OUTPUT? NOT EVEN ONCE?
o4 : Ideal of R
```

The generators of I are actually binomial in the ring where the variables are the marginals, as seen in Equation (1). **....also perhaps just show one of the generators not necessarily all of them, to save space! and even perhaps a couple of the new variables, not all??**

```
i6 : F = marginMap(1,R) -- this is the map where the first variable is marginalized
o6 = map(R,R,{p        - p       , p        - p       , p        - p       ,
               1,1,1,1    2,1,1,1   1,1,1,2    2,1,1,2   1,1,2,1    2,1,2,1
              p        - p       , p        - p       , p        - p       ,
               1,1,2,2    2,1,2,2   1,2,1,1    2,2,1,1   1,2,1,2    2,2,1,2
              p        - p       , p        - p       , p       , p       ,
               1,2,2,1    2,2,2,1   1,2,2,2    2,2,2,2   2,1,1,1   2,1,1,2
              p       , p       , p       , p       , p       , p       })
               2,1,2,1   2,1,2,2   2,2,1,1   2,2,1,2   2,2,2,1   2,2,2,2
o6 : RingMap R <--- R
i7 : I = F I;
o7 : Ideal of R
i8 : netList pack(2,I_*)
     +-------------------------------------+-------------------------------------+
o8 = |- p       p        + p       p       |- p       p        + p       p       |
     |   1,1,1,2 2,1,1,1    1,1,1,1 2,1,1,2|   1,1,2,2 2,1,2,1    1,1,2,1 2,1,2,2|
     +-------------------------------------+-------------------------------------+
     |- p       p        + p       p       |- p       p        + p       p       |
     |   1,2,1,2 2,2,1,1    1,2,1,1 2,2,1,2|   1,2,2,2 2,2,2,1    1,2,2,1 2,2,2,2|
     +-------------------------------------+-------------------------------------+
     |- p       p        + p       p       |- p       p        + p       p       |
     |   1,1,2,1 1,2,1,1    1,1,1,1 1,2,2,1|   1,1,2,2 1,2,1,2    1,1,1,2 1,2,2,2|
     +-------------------------------------+-------------------------------------+
```

This ideal has 5 primary components. The first component is the one that has statistical significance. It is the defining ideal of the variety parameterized by the the factorization of the probability distributions according to the graph G. The remaining components lie on the boundary of the simplex and are still poorly understood.

```
i9 : netList primaryDecomposition I
```

TO DO: we need to decide which of the following examples to include, considering the space restriction. (Also note our margins are larger then those of the journal.) I suggest including either $\{markovIdeal{+}markovRing\}$ **or** $\{gaussianIdeal{+}gaussianOtherFunctions\}$**. the latter should include a reference on Seth's Trek ideal theorem and an example. Then, I suggest running one *identifyParameters* and citing `graphicalmodels.info`.**

4 Technical discussion???

This package requires Graphs.m2, as a consequence it can do computations with graphs whose vertices are not necessarily labeled by integers. This could potentially create some confusion about what does $p_{i_1 i_2 \dots i_n}$ mean. The package orders the vertices lexicographically, so $p_{i_1 i_2 \dots i_n} = p(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n)$ where the labels $X_1, X_2, \dots, X_n$ have been ordered lexicographically. Therefore, the user is encouraged to label the vertices in a consistent way (all numbers, or all letters, etc). how to find the internal sorting?

-MixedGraphs
-trekSeparation
-trekIdeal
-parameter identification
-GraphicalModels.info?

Acknowledgements

Dan Grayson and Amelia Taylor.

References

[1] Gabriel C. G. de Abreu, Rodrigo Labouriau, David Edwards, High-dimensional Graphical Model Search with gRapHD R Package, arXiv:0909.1234v4 **CITE ME!!!**

[2] M. Drton, B. Sturmfels and S. Sullivant, (2009) *Lectures on Algebraic Statistics*, Birkhauser

[3] I. Onur Filiz, Xin Guo, Jason Morton, Bernd Sturmfels, Graphical models for correlated defaults, arXiv:0809.1393v1 **CITE ME!!!**

[4] L. D. Garcia, M. Stillman and B. Sturmfels: Algebraic geometry of Bayesian networks, *J. Symbolic Comput.* **39** (2005) 331–355.

[5] Dan Geiger, Christopher Meek, Bernd Sturmfels, On the toric algebra of graphical models, arXiv:math/0608054v1 **CITE ME!!!**

[6] Hisayuki Hara, Satoshi Aoki, Akimichi Takemura, Minimal and minimal invariant Markov bases of decomposable models for contingency tables, Bernoulli 2010, Vol. 16, No. 1, 208-233 (preprint FOR OUR REFERENCE: arXiv:math/0701429v3) **CITE ME?**

[7] S. Lauritzen, (1996) *Graphical models*, Oxford University Press

[8] E. Stanghellini, B. Vantaggi, On the identification of discrete graphical models with hidden nodes, arXiv:1009.4279v1 **DO WE NEED THIS REFERENCE?**

[9] S. Sullivant: Algebraic geometry of Gaussian Bayesian networks, *Adv. in Appl. Math.* **40** (2008) 482–513.

[10] S. Sullivant, K. Talaska, and J. Draisma: Trek separation for Gaussian graphical models, *Ann. Statist.* **38** (2010) 1665–1685.

[11] L. D. Garcia-Puente, S. Spielvogel and S. Sullivant: Identifying causal effects with computer algebra, to appear in *Proceedings of the 26th Conference of Uncertainty in Artificial Intelligence.*